Reliability Management and Oversight

Reliability Management and Oversight

Brendan Kirby
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Eric Hirst
Consulting in Electric-Industry Restructuring
Oak Ridge, Tennessee

Introduction

Ensuring the reliability of the U.S. power system requires addressing both the system's physical characteristics and the commercial and regulatory frameworks within which it operates. Determining who sets reliability rules and how may be one of the most challenging aspects of maintaining reliability in an increasingly competitive electricity industry. The challenge arises because bulk-power reliability and commerce are tightly integrated; it is necessary for all involved (government policy makers, regulators, consumers, independent generators, and power marketers as well as the utilities that traditionally set and implemented the rules) to understand how bulk-power systems are planned and operated, under both normal and contingency[1] conditions, to participate effectively in commercial markets. And the reverse is true. One should not set reliability standards without understanding how they will affect markets.

The flow of power through the nation's electricity systems is governed by the laws of physics, so an action in one place on the transmission grid affects the entire grid. Thus, although combining individual utility systems into an integrated network increases reliability (by providing redundancy) and saves money (by permitting commerce among regions), interconnections also increase the potential for large-scale blackouts. Because the network is a community asset, its users must cooperate to ensure that it remains viable. And because large-scale blackouts are so onerous, common practice is to take extensive preventive action to assure that they do not occur. This prevention is usually successful, so the impact of reliability issues on the transmission system tends to be economic (i.e., commercial transactions are curtailed and/or power prices are

[1]A contingency is the sudden unexpected failure of a generator, a transmission line, or other piece of equipment connected to the electrical system.

raised in order to maintain reliability) rather than physical (power outages are usually localized and not widespread). Contingency reserves (generation or load that can respond rapidly to system-operator commands and extra transmission capacity that instantaneously accommodates the changed flow patterns) provide reliability by functioning as insurance against the sudden loss of a generator or transmission line.

Managing reliability raises important commercial and societal issues. Reliability rules can favor some commercial entities and exclude others, and these rules affect all of society because they affect electricity prices, electricity availability, and the environment (i.e., the locations of transmission facilities and the amounts, locations, and types of generation, which all have a role in assuring system reliability, have environmental impacts). All users of the power system have an interest in how reliable the system is, what the costs of reliability are, and how decisions concerning reliability are made. Deciding who participates in decisions concerning setting and implementing reliability standards and how consensus is reached involves considerations that range from broad policies regarding the overall level of desired reliability to specific rules governing what is required to participate in particular power system functions. Often the small group that decides specific details of system reliability rules determines the level of risk to which the overall community is subjected. Customers have some choice about the reliability of their electricity supplies—those who require greater reliability than the system provides can install additional equipment (uninterruptible power supplies or individual generators, for example) at their own expense to meet their specific needs. Customers may also have some limited ability to lower their electricity costs by agreeing to accept less reliable service (i.e., by buying interruptible power) or by selling reliability services back to the power system.

The North American bulk-power system is geographically vast, covering the lower 48 states, Canada, and parts of Mexico. It is also organizationally vast, encompassing a wide range of large and small public and private entities, generators, power marketers, transmission owners, transmission operators, and consumers. It must operate in real time within numerous physical constraints, and there are differences of opinion about how best to proceed given these constraints. Mechanisms are needed to make decisions and resolve disputes, which requires authority derived from some established source. This authority could be governmental—federal or state—or based on contractual arrangements.

One organization that has been suggested to administer reliability is the North American Electric Reliability Council (NERC), which is in the process of evolving from a bottom-up, industry-dominated, volunteer organization into the North American Electric Reliability Organization (NAERO), with an independent board. NAERO proposes to set and enforce mandatory standards with regional reliability councils that report to it (rather than vice versa). Regional reliability authorities have also been proposed; these authorities would be free to establish standards that focus on regional conditions.

The requirements of reliability management and oversight must be delineated in order to assess the extent to which alternative institutional structures can meet them. Federally derived authority is attractive because it would provide uniform coverage across the nation, so it would not require the negotiation of numerous parallel agreements.

In the remaining sections, this paper examines the following key issues related to transmission system reliability management and oversight:

- The historical approach to reliability in the U.S., i.e., the creation of control areas and interconnections and the formation of NERC.

- The unique features of the electric power system that affect reliability.
- Reliability from a risk perspective: who causes risk, who is exposed to it, and who pays for reliability.
- The need for and progress toward measuring, paying for, and enforcing reliability.
- Governance issues for new reliability organizations in a restructured electric utility industry.
- Actions that DOE, the Federal Energy Regulatory Commission (FERC), and others should take to improve the reliability of the bulk-power system and our key findings.

Background

Although we know when the lights are off, bulk-power system reliability cannot be easily or unambiguously defined. A reliable electricity system permits few outages or interruptions of service to customers; outages can be defined in terms of their number, frequency, duration, and the amount of load (or number of customers) affected. Equally important, but much more difficult to quantify, is the value of loss of load. A 10-minute power outage in a residence is an annoyance but usually imposes only small economic costs. A similar outage for a computer-chip manufacturer might entail the loss of millions of dollars of output.

Although generation and transmission failures cause only a small fraction of U.S. power outages, their economic and societal consequences can be much greater than those associated with distribution outages. Distribution outages account for the vast majority of customer outage events and outage time. Bulk-power outages, however, generally affect many more customers simultaneously and are much more difficult to recover from than distribution outages. For example, the bulk-power outages that occurred in the western U.S. during the summer of 1996 affected a much larger area and many more people than did the Chicago and New York distribution system outages during the summer of 1999.

The transmission system operator has two basic mechanisms to assure reliability: control of commerce and deployment of reserves.[2] When reliability is threatened, the first mechanism, control of commerce, redispatches generation away from the least-cost (in a traditional, vertically integrated utility) or free-market (in a restructured environment) pattern. This redispatch can be accomplished by means of a number of mechanisms, such as NERC's Transmission Loading Relief (TLR) protocols, reliability-must-run contracts, or locational marginal prices, and is the subject of the paper *Transmission System Operation and Interconnection* by Alvarado and Oren in this volume. The second approach for responding to reliability threats, deployment of reserves, is the primary subject of this paper. Reserves, which can be procured through markets, fit into the categories of extra generation, extra transmission, and load that is willing to curtail in the event of a sudden unexpected failure of generation or transmission.

[2]These two mechanisms are not completely independent. When reserves are acquired, they are taken out of commerce, raising the price of electricity.

Interconnections and Control Areas

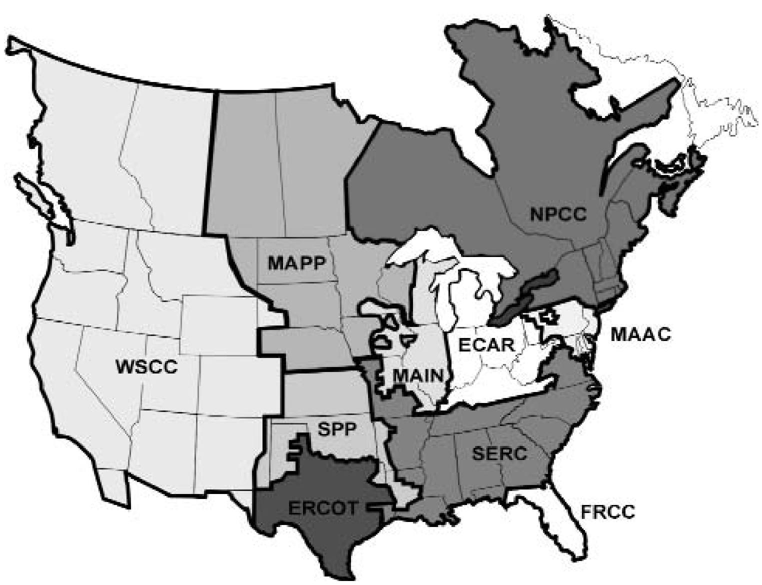
The North American electricity system is divided into three interconnections (Figure 1): the Eastern, the Western, and the Electric Reliability Council of Texas (ERCOT, which covers most of Texas). Within each interconnection, all the generators operate at the same frequency as essentially one machine; generators are connected to each other and to loads primarily by alternating current (AC) lines. The interconnections are connected to each other by a few direct current (DC) links. Because these DC connections are limited, the flows of electricity and trade are much greater within each interconnection than between interconnections.

The entity fundamentally responsible for maintaining bulk-power reliability is the control area. NERC defines a control area as: "An electric system or systems, bounded by interconnection metering and telemetry, capable of controlling generation to maintain its interchange schedule with other Control Areas and contributing to frequency regulation of the Interconnection" (NERC 2001a). Control areas are linked to one another to form interconnections. Each control area seeks to minimize any adverse effect it might have on other control areas within the interconnection by (1) matching its schedules with those of other control areas (i.e., matching generation plus net incoming scheduled flows to loads) and (2) helping the interconnection to maintain frequency at its scheduled value (nominally 60 Hz).

Today approximately 150 control areas are operated primarily by utilities although a few are run by independent system operators (ISOs). Control areas vary enormously in size, with several managing less than 100 MW of generation and the Pennsylvania-New Jersey-Maryland Interconnection (PJM), California ISO, and ERCOT each managing about 50,000 MW of generation. Control areas are grouped into regional reliability councils, of which there are 10 in North America. These reliability regions, in turn, are part of the three interconnections.

The number of control areas and their sizes were set historically; each is the result of the specific manner in which its particular area developed. Although it is likely that there will be fewer and larger control areas in the future, determining the "correct" number and sizes of these areas relative to today's electric power system is a complex combination of technical, political, and market considerations. When control areas are too small and too numerous, coordination among them is difficult. When a control area is too large, it is difficult for the system operator to manage. Unfortunately, some control areas view their autonomy as an economic advantage that they are reluctant to give up; some generators have also sought to

Figure 1: NERC's 10 regional councils cover the 48 contiguous states, most of Canada, and a portion of Mexico.



become autonomous control areas for similar reasons. These attitudes draw attention to the need for true independence of system operators and clearly defined operating rules so that all parties have confidence that they are being treated fairly.

NERC

Historically, the vertically integrated utility industry utilized the North American Electric Reliability Council (NERC) a bottom-up, electric-utility-dominated, volunteer organization to establish reliability rules and monitor compliance. NERC was formed in 1968 in the aftermath of the 1965 Northeast Blackout and in response to the 1967 U.S. Federal Power Commission report on that blackout recommending the formation of an industry-based national reliability organization. NERC is funded by 10 regional councils, which adapt NERC rules to meet the needs of their regions. In 1994 the regional councils opened their membership to independent power producers, power marketers, and electricity brokers, and in 1996 NERC opened its board and committees to voting participation by independent power producers and power marketers (NERC 2001a). NERC and the regional councils have largely succeeded in maintaining a high degree of transmission-grid reliability throughout North America. However, the organization is dominated by representatives of the supply side (generation and transmission) even though the organization's purpose is insure the reliability of supply to the consumer. NERC replaced its 47-member combined stakeholder/independent board with a 10-member independent board in March, 2001. Members of the independent board are still selected by a stakeholder committee, however, rather than being appointed or elected through a political process as regulators typically are.[3]

Historically, the reliability councils have functioned without external enforcement powers and have depended on voluntary compliance with standards. NERC is now in the process of converting to a system of mandatory compliance under which violations will be subject to penalties (including fines). A pilot compliance program is underway to test proposed self-evaluation, data-reporting, and auditing procedures. In the absence of federal legislation requiring compliance with reliability standards, NERC has limited ability to enforce its reliability rules; in case federally derived authority is not forthcoming, NERC and the regional reliability councils are going forward with plans to enforce compliance through contracts and agreements.

Many Western Systems Coordinating Council (WSCC) members have voluntarily entered into contracts committing them to abide by WSCC reliability rules. WSCC is able to impose fines on these members if they fail to meet reliability standards. In this case, contract law, rather than federal regulatory authority, enforces reliability. The severity of sanctions increases with seriousness and number of infractions. However, this is a voluntary process, and not all WSCC members have agreed to these contractual obligations.

---

[3]The NERC stakeholder committee has 35 voting members: one from each of the 10 regional councils, two from investor-owned utilities, two from state/municipal utilities, two from cooperative utilities, two from federal utilities/power marketing administrations, two from merchant generators, two from electricity marketers, two from large end-use customers, two from small end-use customers, two from transmission-dependent utilities, two from ISOs/RTOs, two from Canada at large, and one from Western Canada. There are also six non-voting government representatives: one from the U.S. government, one from each of the three Interconnections, one from the Canadian federal government and one from the Canadian provincial governments (NERC 2001b).

Until a few years ago, FERC and NERC operated on parallel tracks; FERC oversaw bulk-power commerce, NERC oversaw bulk-power reliability, and little interaction was needed between the two. Unbundling generation from transmission and creating competitive markets for electricity have dramatically changed this situation. The industry now recognizes that reliability and commerce are tightly integrated. Increasingly, FERC receives cases in which market participants complain that they face a competitive disadvantage because of NERC reliability rules, their implementation, or both. Partly to address these concerns, and recognizing the growing interaction between reliability and commerce, NERC established a Market Interface Committee as a complement to its long-standing Operating and Planning Committees in September of 1998.

In response to recent NERC requirements, Regional Security Coordinators address reliability issues within the reliability regions and across regional boundaries. These coordinators conduct day-ahead security analysis, analyze current-day operating conditions, and implement NERC's TLR procedures to mitigate transmission overloads.

Reliability Requirements

The electric power system is a communal resource. All users (generators and customers/loads) share the benefits of interconnected system operation. Reliability rules are required to assure that the activities of one user or control area do not adversely impact system reliability for other users or control areas.

Reliability rules require that control areas maintain a balance between generation and load and that they help maintain interconnection frequency. NERC's Control Performance Standards 1 and 2 (CPS 1 and 2) establish requirements for maintaining generation and load balance under normal conditions. The Disturbance Control Standard (DCS) requires that control areas re-establish the generation-to-load balance within 15 minutes of the unexpected failure of a generator or transmission line. NERC also requires voltages to be maintained throughout the power system under normal and contingency conditions. For this purpose, NERC requires that control areas have reserves (extra generation, extra transmission capacity, and/or responsive load) ready to respond immediately when the need arises. These reserves can be obtained through markets, but they must be responsive to system operator commands.

Unique Features of Electricity

Bulk-power systems are fundamentally different from other large infrastructure systems, such as air-traffic control centers, natural-gas pipelines, and long-distance telephone networks. Electric power systems have two unique characteristics:

- The need for continuous and near-instantaneous balancing of generation and load, consistent with transmission-network constraints: this requires metering, computing, telecommunications, and control equipment to monitor loads, generation, and the transmission system and to adjust generation output to match load.

- The primarily passive character of the transmission network, which has few "control valves" or "booster pumps" to regulate electric power flows on individual lines: control actions are limited primarily to adjusting generation output and to opening and closing switches to remove transmission lines from or add them to service.

- These two unique characteristics have four consequences for system reliability, with practical implications that dominate power-system design and operations:
  - Every action can affect all other activities on the grid. Therefore, the operations of all bulk-power participants must be coordinated.
  - Cascading problems that increase in severity are extremely serious. Failure of a single element of the system can, if not managed properly, cause the subsequent rapid failure of many additional elements, disrupting the entire transmission system.
  - The need to be ready for the next contingency dominates the design and operation of bulk-power systems to a greater degree than do current conditions. It is usually not the present flow through a line or transformer that limits allowable power transfers but the flow that would occur if another element failed.
  - Because electricity flows at the speed of light, maintaining reliability often requires that actions be taken instantaneously (within fractions of a second), which necessitates automatic computing, communication, and control actions.

Reliability Functions

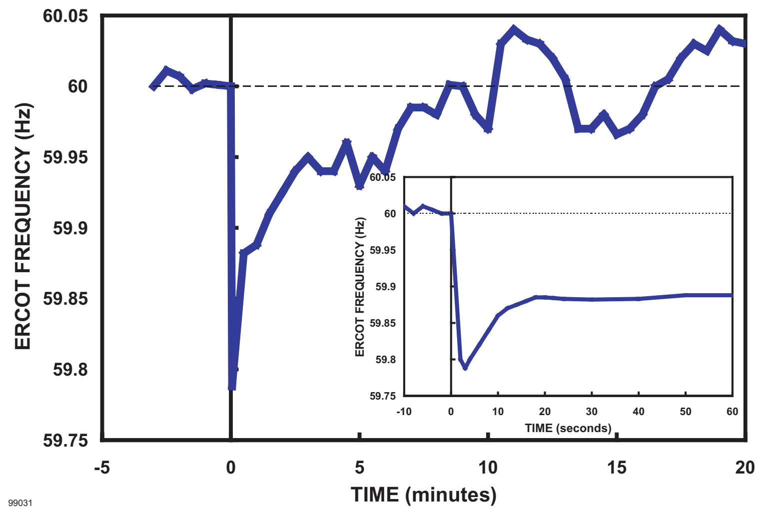
To maintain reliability, the system operator must continuously balance generation and load, maintain acceptable voltages throughout the system, and avoid overloading transmission lines and transformers.[4] Transmission line flows cannot, in most cases, be controlled directly, so line loads must be controlled by placing lines in and out of service and by determining which generators are allowed/required to operate in response to changing load patterns.[5] The interaction between reliability requirements and requirements that determine which generators can/must operate are primarily economic (they restrict transactions and raise prices as discussed by Alvarado and Oren, *Transmission System Operation and Interconnection* in this volume).

It is not sufficient, however, to operate the power system so that generation matches load, voltages are acceptable, and none of the transmission lines is overloaded at the present moment. The power system operator must also be concerned about contingencies—how the system will respond if a transmission line or a generator suddenly fails. Figure 2 illustrates how the electric power system operates when a major generating

[4]The power system is vulnerable to the overloading or sudden unexpected failure of any element of the transmission system. Transmission lines, transformers, circuit breakers, inductors, etc. are all of concern. The term "transmission line" in this discussion refers to all of these elements.

[5]Controlling loads is equally effective but generally harder to do.

Figure 2: Interconnection frequency before and after the loss of a 653-MW generator. The inset shows frequency for the first minute after the outage, and the larger figure shows frequency for the first 20 minutes after the outage.



unit suddenly fails.[6] Prior to an outage, system frequency is very close to its 60-Hz (cycles per second) reference value. Generally within a second after an outage occurs, frequency drops, in this example to 59.79 Hz. The frequency decline is arrested primarily because many electrical loads (such as motors) vary with system frequency and inherent generator response. Frequency is restored through a combination of autonomous (automatic) generator-governor responses and system-operator-directed responses from the generating units that provide contingency reserves. In the example in Figure 2, the system worked as it was intended to, and frequency was restored to its precontingency 60-Hz reference value within the required 15 minutes (at 8.5 minutes in this case).[7] Dedicated contingency reserves are required because there is insufficient time after a contingency to arrange for them. Similarly, there must be sufficient extra capacity available on transmission lines to accommodate the changed pattern of generation that results when contingency reserve generators instantly replace a failed generator. Additional transmission capacity alone may be adequate to accommodate unexpected failures of transmission lines. Alternatively, generation reserves that are closer to the load than the primary generation source can protect against transmission and generation failures.

The overall goal of reliability rules and procedures is to keep customers' lights on. Reliability can be divided into two basic elements: adequacy and security (the accompanying text box gives NERC's definitions for these two terms). Adequacy focuses primarily on assuring that there are sufficient generation and transmission resources available to serve the expected load. Security focuses on the ability of the power system itself to withstand inevitable contingencies. Both concepts involve planning and operations, but adequacy focuses more on planning to assure that enough resources are available, and security focuses more on operations that will permit the power system to remain viable even when unexpected events occur. It is difficult for operators to take actions that restore adequacy if insufficient generation has been built to serve the actual load. Conversely, an inadequate system can still be run securely if the system operator takes actions (which, unfortunately, may include intentional shedding of load) to ensure security.

[6]ERCOT is deliberately used in this example because it is a small Interconnection, so frequency swings are more pronounced there than in the larger Interconnections. It would take an 8,000 MW drop in generation in the Eastern Interconnection to obtain the same frequency drop as in this example, and no generating units are that large. In this regard, larger interconnections are "better" than smaller ones because more generators are available to respond to emergencies; however, there also has to be enough transmission capacity to adequately couple the generators and to keep the system stable.

[7]At the time of this disturbance, NERC's allowable disturbance-recovery period was 10 minutes.

NERC's Definition of Reliability

NERC, the primary guardian of bulk-power reliability, was established in 1968. NERC's creation was a direct consequence of the 1965 blackout that left almost 30 million people in the northeastern United States and Ontario, Canada, without electricity.

NERC defines reliability as "the degree to which the performance of the elements of [the electrical] system results in power being delivered to consumers within accepted standards and in the amount desired." NERC's definition encompasses two concepts: adequacy and security. Adequacy is defined as "the ability of the system to supply the aggregate electric power and energy requirements of the consumers at all times."Security is defined as "the ability of the system to withstand sudden disturbances."

In plain language, "adequacy" implies that there are sufficient generation and transmission resources available to meet projected needs plus reserves for contingencies. "Security" implies that the system will remain intact even after outages or equipment failures.

Load as a Reliability Resource

The inherent responsiveness of loads to power system conditions and the system operator's (limited) ability to control loads both have important implications for reliability. Motor loads inherently reduce their power demand as system frequency falls, for example, helping to stabilize the power system when generation is lost. Similarly, heaters and incandescent lamps reduce their power consumption when voltage drops. This "natural" response has diminished in recent years as solid-state power-conditioning equipment compensates for changes in delivered power; for example, as voltage or frequency drops, load-controlling equipment increases the consumption of current to maintain the energy being delivered to the load. There are some benefits to solid-state load control, however. Some load control equipment is designed to disconnect loads to protect against damaging undervoltage. This response prevented a voltage collapse in a major U.S. city recently even though the response was uncoordinated and unplanned.

The system operator can also control how much load is served. Some loads respond to price signals; other loads may be directly under the operator's control (i.e., the customer has agreed to have load curtailed at the system operator's discretion). Load control, especially based on customer response to market signals, is an underutilized resource for helping ensure system reliability (Hirst and Kirby 2001a). Loads that respond to energy price signals tend to mitigate reliability problems because energy prices are often high when the power system is stressed and generation resources are scarce. Customers who defer energy consumption to time periods when prices are lower help themselves by reducing their energy costs, help other customers by reducing energy price spikes, and generally increase system reliability by improving the generation/load balance. Loads that specifically sell reliability reserves to the power system (currently a small number) are treated in the same fashion as generation reliability reserves; that is, they improve reliability by increasing the reserve supplies.

Operators also have the crude ability to "control" (disconnect) loads that have not agreed in advance to be curtailed. When the power system is under severe stress, the system operator's primary focus shifts from providing all loads with electric power to ensuring the system's viability. In the worst case, some loads become a resource whose primary function is to stabilize the power system. The system operator uses the only control over loads that is generally available: deliberately disconnecting blocks of loads. In this situation, system security is maintained at the expense of adequacy. Although this might at first appear to be a conflict in priorities, it is not. Serving loads safely, reliably, and economically is still the system's priority. Under these unusual conditions, however, load can best be served by securing system viability first and attending to loads second. This approach is preferable to the more difficult and lengthy process of restoring the power system after a major regional collapse, which disrupts service to all customers in the region. Curtailing service to a few customers to maintain the system's viability greatly reduces the total number of customers who are affected. Deliberate curtailments also generally leave customers without power for shorter periods than would be the case during a regional outage. Intentional curtailments can result from automatic relay action (under-frequency or under-voltage load shedding) during a disturbance. They can also result from system operator action, either preemptive—as with California's rolling blackouts in 2001—or in response to a disturbance.

Although load curtailment events are rare, they are important as the last line of defense before the power system collapses. Preparing for them requires considerable planning. Agreeing to the rules under which they are implemented requires consensus among technical, business, and regulatory interests. Rules governing how the system operator uses involuntary load curtailment should be publicly established and available.

Risk

Power system reliability management is risk management—a tradeoff between lower costs and greater reliability. The communal nature of the transmission system means that all users share the risk. The fundamental reliability management and oversight issues are determining what risks to take, when to take them, how much money to spend on risk mitigation, who pays for reliability, who is exposed to the remaining risks, and who decides on these matters. These questions are much more complex since restructuring than they were for the vertically integrated industry of the past. Finding satisfactory answers requires obtaining consensus among technical, business, and regulatory interests.[8]

The vertically integrated utilities of the past and their regulators implicitly agreed on the level of reliability to be maintained (and, therefore, on the amount of generation and transmission reserves that each utility carried). Greater flexibility existed for responding to changing risks. A system operator of a vertically integrated utility, for example, could decide to decrease dependence upon long transmission lines when a thunderstorm approached the service area by reducing remote generation and increasing generation close to the load. The increased cost of the off-economic dispatch was borne by all customers if the regulator approved of the practice. The key cost was the differential in production costs between cheap remote generators and expensive local generators. Customers saw this cost only as a slight increase in their average annual rates. Little analysis may have been required to justify this practice – and little analysis might have been possible because of the difficulty of precisely quantifying the change in outage probability or the cost of outages. Thus, implementation might have been left to the judgment of the system operator. This practice of altering

[8]This consensus must respect the laws of physics, but there are generally multiple ways to address any requirement, and technical concerns are not the only, nor necessarily the dominant, ones to address.

reliability rules based on the system operator's judgment and experience may not be permitted in the restructured industry, however. In particular, independent remote generators would object to their sales being curtailed. Impacts on remote generators are much greater than the simple production cost differential between remote and local generators; for independent generators, the difference is between running at a profit and siting idle at a loss. In addition, the remote generator might have to pay for the operation of the competing local generator if the remote generator had made a firm sale to local customers. Remote generators will insist on detailed analysis to demonstrate that the curtailment of their facilities was necessary, beneficial to the system, and done in a nondiscriminatory way. It is important for the system operator not only to be independent of commercial concerns but to be perceived as being independent. To the greatest extent possible, the system operator's decisions should be based on analysis rather than personal judgment. The analysis methods and results should be made public along with data concerning system performance.

The view from the customer's or load's perspective is somewhat different. It is primarily the loads that are vulnerable to the risk of system failures and blackouts. It is also the loads that pay the higher costs associated with greater reliability. In the future, customers may want to participate more directly and fully in the rule-making process, along with the traditional participants (generation and transmission companies).

Similarly, society as a whole and the governmental bodies that represent and protect it have an interest in power system reliability. While local power outages primarily affect customers in the immediate area, widespread outages have a disproportionately larger impact. Public safety is threatened. Police and fire departments can be overwhelmed with response calls. All commercial activity halts in the blacked-out region. These negative consequences of outages are the reason that the power industry has historically emphasized system security at the expense of reliability for individual customers even though the purpose of the power system is to deliver reliable power to customers.

Adequacy and Security

As noted above, adequacy focuses on ensuring, in the long term, that sufficient generation and transmission are planned, designed, built, and available to meet load requirements. Security addresses the short-term survival of the power system when disturbances occur. These two characteristics of reliability interact. A system with ample generation and transmission resources will be adequate and (if run well) secure because there will be sufficient resources to serve load and respond to contingencies. Adequacy, security, or both are reduced when there are not enough resources to serve all load requirements with sufficient additional reserves to address contingencies.

Adequacy can be maintained at the expense of security. That is, the power system can serve its full load without holding back reserves, but the resulting risk is that it would not survive a severe contingency. The risk period may be limited to those few hours per decade when loads are particularly high or when generation or transmission equipment is out for maintenance, or the risk may be much greater and more frequent if the system is seriously deficient in resources. Risk probability differs at different times as well, e.g., transmission line outages are much more likely during a thunderstorm than on a clear day.

Just as adequacy can be maintained at the expense of security, security can be maintained at the expense of adequacy. That is, load can be curtailed to maintain generation and transmission reserves that protect the system against contingencies. Load can be controlled by means of market mechanisms (responsive load programs) or through command-and-control mechanisms (rolling blackouts or underfrequency relays). Curtailments may be limited to small geographic areas or they may be system wide, depending on need. A curtailment may be necessary for only a few hours every decade or as often as daily.

All power systems balance adequacy and security in addressing reliability. It is not practical to build a power system that can withstand all contingencies or that will remain adequate under all circumstances. Because the system cannot be 100 percent reliable in practice, the important questions are who takes what risks when and who decides on the rules.

Risk Response

Contingency reserves and operating rules that govern their use are the primary mechanisms to mitigate risk. A brief look at the function of contingency reserves gives insight into how reserve rules have been established and may help to guide how reserve rules should be set in the future. Contingency reserves are resources that are kept out of service in anticipation of the sudden failure of a generator or transmission line. Their purpose is to address a probabilistic problem (the statistical event of a contingency occurring); however, contingency planning has been treated as essentially deterministic. That is, the NERC requirement for ensuring bulk-power-system reliability is deterministic in that it requires that the power system be continuously able to withstand any single contingency regardless of the probability of occurrence, the cost to protect against it, or the cost of failing to protect against it. Specifically, NERC requires all control areas to operate so that "instability, uncontrolled separation, or cascading outages will not occur as a result of the most severe single contingency. Multiple outages of a credible nature shall also be examined and, when practical, the CONTROL AREAS shall operate to protect against instability, uncontrolled separation, or cascading outages resulting from these multiple outages." (NERC 2001a)

There is a probabilistic nature to deciding which *multiple* contingencies are credible and should be considered; system planners and operators use an informal, deliberate, closed process to decide which contingencies are credible and which are not, and what types of events the system should be designed to survive. The loss of any single generator or line (the N-1 criterion), for example, is almost always considered. The simultaneous loss of both circuits in double-circuit configurations is also often considered. The simultaneous loss of multiple generators at a single generating plant may be considered if there are common-mode failures that can affect multiple generators.[9] Decisions about what to take into account are primarily based on the planners' and operators' experience with the power system rather than on detailed probabilistic calculations.

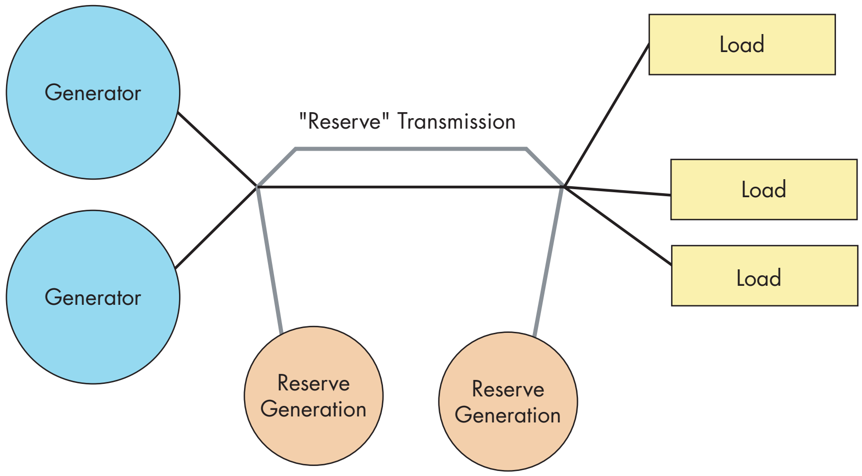
Once the process of deciding which contingencies merit the expense entailed in guarding against them and which contingencies are sufficiently unlikely that they do not, the process becomes more deterministic.

A simple example system

Figure 3 presents a simple isolated power system consisting of two generators supplying loads through a single transmission line. The reserve requirements are straightforward. The generator output and line flow are always equal to the total load. Contingency reserves equal to the current generator output are required con-

[9]A "common-mode" failure is a single event that could trigger what would otherwise be considered a multiple contingency. The failure of a common cooling water supply or a common fuel supply could cause the simultaneous failure of multiple generating units.

Figure 3: Contingency reserves compensate for the unexpected loss of generation or transmission in this isolated example system. Increasing the number of generators and reducing their individual size reduces the required generation contingency reserve.



tinuously. The contingency reserves can be made up of any combination of generation that is able to come on line quickly enough and load that can be removed quickly enough.

It might be rational to decide that the probability that both generators will fail at the same time is low enough that it is cost effective to provide only enough reserves to cover the loss of a single generator at a time. That decision would cut the reserve requirements in half (assuming the two generators are of equal size). The generators themselves can supply the reserve for each other. For example, each could be loaded to half capacity; the unloaded half capacity remains available as the reserve for the other generator. If there are 10 generators instead of two, it is easier to see the attractiveness of this option, with the reserve requirement being equal to the largest amount of load being carried by any one generator. Reserve requirements remain high (equal to the total generation rather than individual generating unit output), however, if transmission contingencies are also considered because both generators use the same transmission line. In any case, determination of the reserve requirements is a deterministic process once the credible contingencies have been selected.

The N-1 criteria appears straightforward in this case, though there are some complexities. Failure of either of the generators or the transmission line will affect all of the loads; reserves must protect against both contingencies. This requirement means that reserve generation must be located near the loads or a combination of remote reserve generation and reserve transmission must be maintained.[10] The reserve requirements now become deterministic. It does not matter whether the generator typically fails once an hour, once a week, or once a year. The reserve requirements remain the same until the probability of failure becomes low enough that the reserve requirement can be eliminated altogether. The failure rate may influence the choice of facili-

[10]For simplicity, the reserve transmission is shown as a separate line in the figure. Because it is not economic to directly control flows on individual AC transmission lines within a network, both the “primary” and “reserve” capacity would actually be on the same physical transmission lines. For example, two parallel lines could each be loaded to half of their individual capacities, or three parallel lines could each be loaded to two-thirds of their individual capacities.

ty that provides the reserves but does not affect the number of megawatts that must be maintained.[11] The amount of reserves required is equal to the amount of precontingency generation (ignoring changes in losses). If any less is maintained, the system collapses because post-contingency load will exceed generation.[12] Maintaining any more is a waste of resources.

Multiple versus single contingencies

The simple example above demonstrates how the probability of a single contingency does not affect contingency reserve requirements unless the contingency is so improbable that it can be ignored completely. Interestingly, multiple contingencies also exhibit a similar nonprobabilistic characteristic. If we greatly oversimplify to examine the underlying concept, we can assume a power system with 500 critical transmission lines in which the typical transmission line experiences an unscheduled outage once every 10 years. This means that the system operator faces about one contingency per week. The power system cannot be allowed to collapse on a weekly basis, so the system must be protected against single contingencies even though they each only occur once every 10 years.

The probability of multiple simultaneous contingencies is much lower. Assuming that a typical transmission outage lasts 0.1 hours (most are restored through automatic recloser action in a shorter time; others last longer, but the system operator takes corrective action within a fairly short time to reduce the system's vulnerability), the probability of a second line failure occurring while the first line is out of service, a double contingency, is reduced to one event in 35 years. In addition, many of these double contingencies will be sufficiently separated electrically so that they will not have compounding effects. Assuming that 25 percent of the double contingencies threaten system viability reduces the risk to one event every 140 years. This is an 7,008:1 ratio in the probability of a single versus a double contingency.

Although the simplifying assumptions and numbers in the example above have little relationship to reality, they illustrate an important point: the difference in probability of single versus multiple contingencies is so great that it may be reasonable to ignore multiple contingencies unless there is a common failure mode. This reasoning helps explain how power system planners and operators were able, when the industry was vertically integrated, to independently assess reserve requirements and reliability rules without needing extensive consultation with loads, generators, regulators and others.

Deliberate damage to the power system is an ever-increasing concern for utilities, law enforcement, policy makers, regulators, and the public. Increasing the attention paid to power system reliability in general will help reduce the system's vulnerability to terrorism. Deliberate attacks on the power system pose unique concerns, however, as addressed in the accompanying text box.

---

[11]If the primary generator or transmission line fails infrequently, then the reserve generator should have a low capital cost but may have a high operating cost. If the primary supply fails once a week, it may be better to invest more in the contingency supply to lower operating costs.

[12]This is not strictly true. If the load is highly frequency sensitive, the system could settle to a stable lower frequency if "almost" enough reserves were available.

"Flexibility" in Reserve Requirements

The real world is not quite as straightforward as the simple systems in the example above where reserve requirements appear to be completely deterministic: either the system has enough reserves in the correct places to survive the contingency or it does not; if it does not, the system will collapse in a contingency.

The contingency reserve requirements outlined in the NERC and regional reliability council guidelines are similarly deterministic. Most regional councils establish contingency reserve requirements based on the size of the largest single contingency for the reserve-sharing group. WSCC requires contingency reserves equal to five percent of hydro generation and seven percent of thermal generation. Most require that at least half of the contingency reserve be spinning [NERC (2001a) defines spinning reserve as "unloaded generation that is synchronized and ready to serve additional demand"], but some only require that 25 percent of reserves be spinning (FRCC 1999). Which requirements are "best" and why? We were unable to find analysis documenting the reasons for the reserve requirements in any region. These analyses should be conducted, documented, and made public so they can be assessed by market participants and government regulators.

Determining what constitutes sufficient reserve is complicated by the networked nature of the transmission system, the difficulty of modeling the system exactly, and the dynamic interactions among multiple generators and

Deliberate Damage

The power system is designed and operated to withstand the unexpected failure of any single generator, transmission line, transformer, or other piece of equipment. It is also designed to withstand the simultaneous loss of multiple pieces of equipment if there are known physical reasons why they would fail simultaneously. Transmission lines that share transmission towers, for example, could fail simultaneously if a tower was damaged. This design philosophy provides solid protection against natural threats, such as lightning or falling trees. It also provides protection against some types of deliberate damage, e.g., an individual hunter shooting out transmission line insulators or someone toppling a transmission tower.

However, the power system is not typically designed to withstand the simultaneous failure of multiple pieces of equipment from either natural causes (e.g., hurricanes) or deliberate acts of sabotage. One reason that these types of contingencies are not guarded against is that the networked nature of AC power systems means that (a) it is quite expensive to protect against all multiple contingencies and (b) multiple contingencies are quite unlikely. Furthermore, the typical approach to protection—maintaining reserve transmission and generation capacity—is of limited use against a large-scale threat; a hurricane or saboteur is as likely to damage six transmission lines as to damage two. Determining which multiple contingencies to which the power system is currently vulnerable is technically complex and requires extensive system knowledge.

This means that the nature as well as the amount of protection is different when we consider the risk of deliberate damage. For example, reserve generation located closer to load is of more value in protecting against widespread damage to the transmission system than is additional transmission capacity. Determining what actions should be taken to protect the electric power system from deliberate threat is of great concern to DOE's Critical Infrastructure Protection Program.

loads. All of these factors complicate predictions of post-contingency conditions. Moreover, during a contingency, additional support may be available from other sources or in other forms; for example, a control area can typically get reserve from the interconnection, and a system operator may have resources such as direct control of load shedding that can be manipulated rapidly to restore the generation-load balance. Together these factors make it very difficult to determine the exact reserve requirements that are appropriate for any instant in time.

If a system operator finds that load exceeds expectations and reserves are not available (or are extremely expensive), what can be done? The operator will curtail all nonfirm transactions and seek help from neighbors, etc., but when all options are exhausted and load plus reserve requirements exceed available generation capacity, what should the operator do then? One possibility is to curtail firm load (i.e., black out some customers) in order to preserve reserve margins and avoid risking a regional collapse. Politically, this is very difficult to do (in either the vertically integrated utility environment or in the restructured environment) because blackouts receive national attention. They point out that the system failed to prepare adequately for the events that led to the inadequate supply of energy and reserve, and they frequently prompt a phone call from the state governor to the system operator's chief executive. It is difficult to explain to the public that a system operator deliberately cut off power to customers because there was a chance that a generator or transmission line might fail and cause problems.

Because of the pressure to avoid loss of power to customers, there is a strong temptation to deal with inadequate reserves without curtailing firm load. System operators admit privately that they have commonly drawn down contingency reserves rather than curtail load. This concern is difficult to document in vertically integrated utilities because many of their operating procedures are not published. Industry restructuring and the establishment of ISOs have made reliability rules more specific and more public. The California ISO, for example, does not initiate rolling blackouts until operating reserves fall to 1.5 percent or less, which is well below the WSCC five- to seven-percent reserve requirements (California ISO 2001, WSCC 2001). Similarly, ISO New England provides for operations "which may result in degraded system reliability since the full operating reserve that is required for normal operation is not maintained" before the system operator resorts to intentional load curtailment (ISO New England 2001). Deferring the curtailment of load has consequences, however; it compromises reliability in neighboring control areas and throughout the interconnection.

Two Questions about Community Risk versus Individual Benefit

Two significant, distinct issues in power system contingency response are whether the danger of a regional collapse is increased by reliability decisions, and who pays and who benefits as a result of these decisions.

Reducing reserve margins to the extent that the power system is at increased risk of collapse (or taking any other action that increases the collective risk) has serious consequences for all users of the system and for society as a whole. The loads and society suffer the consequences if things turn out badly. Determining when the power system has moved from one level of risk to another is highly technical. Determining whether the power system should move from one level of risk to another is a commercial, political, and regulatory question that should be debated in a public forum.

Replacing conventional generation reserves with dependence on the interconnection, fast operator action, load response, or other similar strategies for responding to a contingency raises fewer societal concerns than reducing reserve margins, as long as these strategies successfully prevent a system collapse. These strategies raise commercial and regulatory issues for the individuals involved, however.

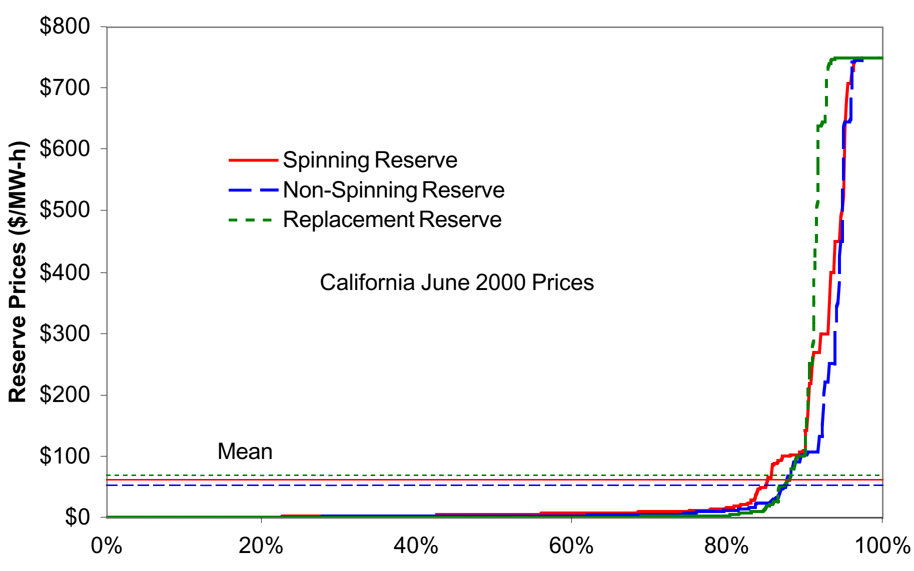
The communal nature of the transmission system means that risk to the system is generally assumed by the community at large. If an "incredible" contingency occurs or if contingency reserves are inadequate, then some or all of the system collapses. All of the loads (customers) connected to the collapsed portion suffer. For a large outage, additional societal costs (traffic, police, etc.) are borne by the affected region. Generators within the collapsed portion of the system and generators selling to loads in the collapsed portion suffer losses as well, but the loads suffer the most significant loss.

An individual control area that finds itself short of generation exposes neighboring control areas and loads to increased risk of system collapse if it uses its own reserves to serve load. It is implicitly relying on reserves in other areas as backup but without paying the other areas for this "service." In a restructured electricity industry, each commercial entity is responsible for contracting for an adequate energy supply. If a load fails to contract for enough supply or if the supply fails to deliver, then using contingency reserves to cover the shortfall simply rewards the poor performer and exposes good performers to added risk. Because the poor performer likely saves money by means of this behavior, it has incentive to repeat its poor performance.

Risk and Price

Historically, some system operators operated (at least occasionally) with reduced reserves when reserves were simply not available. In the restructured environment, should the operator's choice about how much reserves to procure be affected by the price of those reserves? Contingency reserve prices vary dramatically, as shown in Figure 4. Although the price of spinning reserves was typically quite low in June, 2000 (the median price was only $5/MWh), it reached its $750/MWh cap two percent of the time and was above $100/MWh 12 percent of the time. The spinning reserve price was below $15/MWh 80 percent of the time. These statistics indicate that there were many hours during which reserves could have been increased at relatively little cost and a few hours during which reserves could have been decreased in return for relatively large savings.

Figure 4: Contingency reserve prices in California are typically low with occasional price spikes.



The probabilistic nature of contingency reserves argues for adjusting the amount procured based on the price. Deciding which contingencies are credible and must be protected against is basically a probabilistic economic tradeoff.[13] That is, we don't protect against contingencies that are too unlikely and/or whose consequences are very low cost. Said another way, we stop protecting against contingencies when the cost of the protection exceeds the benefit of avoiding the contingency. Unfortunately, it is difficult to rigorously analyze this economic tradeoff. Useful data on contingency probabilities and the costs of the contingencies (the value of lost load) are not available.

It is also difficult to unambiguously value load that is lost when the system operator curtails load either preemptively with rotating blackouts or during a contingency. This load shedding is involuntary for the loads affected but voluntary for the system in that the system operator deliberately uses load to restore or instead of generation reserves. An "advantage" to the system and the loads that are not curtailed is that the loads that are involuntarily shed are generally not compensated. Compensating curtailed loads would reduce the economic incentive to push the cost of involuntary load interruptions onto certain loads chosen by the system operator.

An improved system would be to establish a market-based program for loads to voluntarily offer, for a price, to immediately curtail in the event of a contingency. Such a program would formally recognize voluntary load response as a legitimate contingency reserve. Technology would have to ensure that the response was as fast and accurate as that offered by generators. It is important to note that this use of load as a contingency reserve resource does not change the system's reserve response; it only reduces the need for generation to supply that reserve. ISOs expanded their emergency and economic demand-response programs in 2000 and again in 2001 as an initial step in this direction.

New risks: common-mode failure, gas, trading hubs, and time

Making the amount of reserves carried by the system price-sensitive introduces a potential common-mode failure. In interconnections with multiple reserve-sharing groups, each group is individually responsible for its own reserve requirements. Because frequency is common throughout the interconnection, however, the groups support each other in the event of a major contingency. If one group has underestimated its reserve needs or if its reserves fail to respond adequately, that deficiency will likely be made up by another group. Tying the amount of reserves available to reserve prices will mean that all groups will tend to procure fewer reserves when prices are high, increasing the system's vulnerability to collapse.

Natural gas is an attractive fuel for producing electric power, especially for new generators. The capital costs for gas-fired generating plants are lower than those for coal-fired plants. Natural gas emissions are also inherently lower, so environmental mitigation costs are reduced relative to those for coal plants. Gas-fired plants can be built much more rapidly, often within two years versus seven to 15 years for coal-fired plants. There are often fewer siting problems for gas plants as well. The higher cost of gas is one of the few disadvantages. Gas-fired plants also have reliability benefits. They are typically faster to start and to respond to load-change commands than coal plants. This increases their value in providing contingency reserves. But gas-fired plants also raise reliability concerns primarily related to the inability to store gas. In contrast to coal-fired power

---

[13]ISO New England staff members indicate that the inability to quantify the costs and benefits of contingency protection has stalled efforts in New England to formalize reserve-price response.

plants, which in the past typically had 30- to 60-day fuel inventories on site, a gas shortage hits all plants quickly.[14] Worse, gas-fired generators fed from the same gas pipeline are tightly coupled. Failure of a single gas pipeline may constitute a region's single worst contingency. For example, consider the situation in Arizona where approval has been given for the siting of 12,000 MW of generation, all of which would be served from the same pipeline (this is almost 10 percent of the current WSCC capacity) (Smith 2001).

Trading hubs may present a similar problem. If developers of new generation find it economically attractive to locate close to trading hubs, the resulting increased concentration of generation will result in multiple generators depending on the same transmission path even if the generators themselves are sufficiently independent to have no common-mode failures. An example of such a trading hub is Palo Verde, Arizona where a large number of generators are requesting interconnection.

The long life and high cost of power system equipment present challenges to reliability in a competitive industry. Large transformers, for example, can operate for 40 years or more, but their longevity is tied to how they are operated. Heat (overloading) is a major contributor to insulation degradation and equipment failure although this effect is difficult to quantify as it takes place over time. This fact makes it extremely difficult to put a price on an emergency response action by a competitive generator (or, potentially, a competitive transmission provider) because, although temporarily overloading a piece of equipment may shorten its life by some amount of time, its failure is not likely to be immediate. How should a competitive supplier factor in the impact on equipment when pricing emergency response? And how can a system operator know what a reasonable price is? If the equipment owner does not feel that it is being compensated for its risk (either through overall price or payment for the event), it may not respond to reliability events.

Decision making and risk taking

Decisions about procuring contingency reserves are made by power system planners and operators, but customers, and, to some extent, generators, face the resulting risks. This split between the decision maker and the risk taker was the same when utilities were vertically integrated, but the consequences were not as dramatic because the system operator typically also owned the generation, and the transmission and often the distribution systems, and regulators could hold the company responsible for its overall performance in supplying energy to customers. Although FERC and state regulators will continue to oversee system operations, transmission, and distribution in the restructured environment, the responsibility for overall performance of energy delivery is now split among generators, system operators, transmission owners, etc. However, the customer still pays the consequences of unreliable energy supply. Customers should therefore be involved in determining the amount of risk to which they want to be exposed and how much money they are willing to pay to avoid the risk of widespread blackouts.

With a properly structured market each customer can, to some extent, decide individually what level of reliability s/he is willing to pay for. A customer can select interruptible power if price is more important than reliability, for example, or a customer might decide to sell reserves to the power system if the price is attractive, if it can respond fast enough, and if the reserve supply rules are technology neutral. In a fully functional future energy market, adequacy may be the responsibility of each supplier. If a supplier fails to provide ade-

14Competitive pressures are pushing coal-fired generators to reduce their coal inventories as well. Some generators now maintain inventory of 10 days or fewer.

quate resources and if it cannot obtain them from the market, then that supplier's customers would be curtailed. The market would apply appropriate pressure on suppliers to maintain adequacy.

It is equally important that all parties (loads, generators, regulators, transmission owners, and system operators) be involved in the communal decisions that determine the level of security that the system should maintain.

Measuring Reliability Response

Generator response to system operator commands during contingencies continues to be an area of major uncertainty that is intensified with restructuring. In the vertically integrated utility environment, the system operator that was responsible for reliability belonged to the same corporate entity that owned the generation that provided contingency reserves. Measuring generators' response to contingency orders might have been useful for the internal operations of the utility, but it was not critical for judging the overall performance of the system. Restructuring has separated the system operator from the generation resources and thus created a great need for metrics to assess performance (Hirst and Kirby 2001b). Without metrics, it is difficult to know whether a generator is providing the reserve service that it is obligated to provide. Unfortunately, metrics have not been established in most regions to effectively gauge performance when contingency reserves are called upon. NERC started to develop a policy (Policy 10) on Interconnected Operations Services (essentially FERC's ancillary services), but this effort has stalled, and the policy remains only a reference document. It is likely that performance will deteriorate in the future if there are no clear service definitions and metrics on which to base compensation and nonperformance penalties. Some of the contracts that independent power producers (IPPs) and transmission service providers (TSPs) signed when IPPs bought existing generating units failed to take account of some "ancillary services" that were still needed to support the system, such as reactive supply and voltage control. Because IPPs have no way to get paid for these services in the absence of contracts that address these issues, some producers are reportedly balking at providing the services. Creating markets for reliability services would establish a means for compensating service providers.

WSCC has implemented a Reliability Criteria Agreement to enforce reliability requirements in a restructured environment (WSCC 2001). This agreement parallels NERC's Pilot Compliance Program, but WSCC's contracts with its members allow for enforcement. WSCC assesses compliance based on five criteria for control area operators:

- Operating Reserves—each control area is required to maintain regulating and contingency reserves (spinning and nonspinning).
- Disturbance Control—each control area is required to successfully respond to each contingency (restore area control error within 15 minutes).
- Control Performance Standards 1 and 2—each control area is required to meet NERC control performance standards (limits on area control error under normal conditions).
- Operating Transfer Capability —each control area is required to keep flows over transfer

paths (transmission lines) within the Operating Transfer Capability (OTC) Limits of each transfer path. Stability, thermal, and/or voltage constraints set OTC limits.

WSCC also enforces compliance for generators based on continuous operation of the generator's power system stabilizer and automatic voltage regulator. These criteria are concerned with whether generators are maintaining their fast-response capability to maintain power system stability.

Note that the first criterion above (operating reserves) ensures that the control area continuously maintains the required reserves. The second criterion (disturbance control) addresses whether the reserves actually respond effectively when contingencies occur. The third and fourth criteria (control performance standards 1 and 2) assess whether reserves respond effectively during normal operations. The fifth criterion (operating transfer capability) focuses on whether transmission reserves are continuously maintained.

Penalties for violating reliability criteria range from a letter sent to the violator's chief executive (for an initial violation at a relatively low level) to fines of $10,000 or $10/MW, whichever is higher (for multiple violations at higher levels). Levels are determined by the amount of shortfall relative to the criteria. Allowing operating reserves to dip to between 90 and 100 percent once during a month earns the control area operator a letter to its chief executive. Allowing this shortfall twice during a month or dropping between 80 and 90 percent once in a month typically results in letters to the chief executive and the chairman of the board of the offending party, the state or provincial regulatory agency, FERC, and the U.S. Department of Energy (DOE). Allowing operating reserves to drop to between 70 and 80 percent or repeating earlier infractions starts to cost the control area $1,000 or $1 per MW of shortfall, whichever is greater. Dropping below 70 percent or continuing to repeat earlier transgressions increases the financial penalty.

The penalties for violating the other reliability criteria are identical although the metrics are specific to each criterion. The exception is violations of the disturbance control criteria; the penalty for these violations is an increased contingency reserve criterion for the subsequent three months.

Experience to date is not conclusive regarding the effectiveness of the WSCC Reliability Management System (RMS) system in improving performance. To date, $2.2M in sanctions have been assessed against participants in the RMS program and $0.3M against non-participants (not all WSCC members currently participate in the program) (Dintelman 2001). Performance in some categories (maintaining automatic voltage regulators, for example) seems to be improving for participants but not for nonparticipants, which might be an indicator that the program is effective. However, compliance with the DCS appears to be improving for both participants and nonparticipants while the number of noncompliance incidents for all RMS categories appears to be growing for both groups as well.

None of the systems or proposals that we examined linked the penalty for a control area's or a reserve supplier's nonperformance to the cost consequences. As was true in the past, loads that are curtailed, either proactively by the system operator to manage a contingency or as a direct result of the contingency, are not compensated for their losses. Similarly, no attempt is made to quantify and compensate for societal damages (police and fire response, etc.) that result from a widespread outage.

Perhaps worse, data concerning the number of customers subjected to power failures or unacceptable power quality, the time taken to restore power, and the amount of power not delivered are not publicly available.

This makes it impossible to know whether reliability is improving or declining and to what degree. (These data are commonly available in the United Kingdom and much of Europe.)

Governance

As discussed above, the utility technical community (through the NERC committee structure) historically set reliability management rules. NERC was "owned" by the regional reliability councils which, in turn, were "owned" by the member utilities. This structure is beginning to change as NERC and the regional councils open their membership and boards to nonutility participation. Still, staff from transmission and generation entities dominate the committee structures.[15] Although this imbalance in representation may make sense from a technical standpoint, it leaves customers with little ability to influence the reliability decisions with whose consequences they must live.

The question of governance and independence is problematic for all organizations that attempt to be neutral facilitators. The California ISO, for example, has been criticized for not being sufficiently independent. The *Energy Daily* (Davis, 2001) raises questions about the ISO's dealings with the states in its attempts to obtain sufficient power at reasonable cost for consumers:

> The ISO, critics charge, is far from independent, and its actions could stretch beyond providing information to DWR [the California Department of Water Resources] and include manipulating power prices to prevent the [California] governor from being embarrassed by a huge gap between market prices and prices in the long-term contracts signed earlier this year.

The authority behind reliability rules is also problematic. Because NERC is a voluntary industry organization, it has no enforcement power. For many years, its reliability rules were little more than best practices or guidelines. The real authority came from state regulators who had power over individual utilities. State regulators and FERC tended to defer to NERC on technical matters, so a utility that abided by NERC rules was generally regarded by regulators as behaving prudently. To date, only WSCC has found an alternative method, voluntarily entered binding contracts, to establish reliability authority. NERC and the regional reliability councils are proceeding with plans to enforce compliance with reliability rules through contractual agreements in case congressional action and federally derived authority are not forthcoming.

With the increased commercial activity brought about by restructuring, there is a great need for clearly defined operating and planning rules. The commercial separation of generation from system operations, and of one generator from another, makes for a healthy competitive environment but one in which everyone is

[15]Although NERC has made an effort to open the committee structure, the NERC Roster reveals continued dominance by generation and transmission entities (NERC 2001b). The Operating Committee, for example, has 33 members of which 22 work for public, private, state, and federal utilities, five work for independent system operators, two represent IPPs, two represent power marketers, and two represent customer groups. The Interconnected Operations Services Subcommittee, which is tasked with developing Policy 10 on ancillary services, has 20 members of which 15 work for utilities, two work for independent system operators, two work for power marketers, none represent loads or regulatory interests. These are typical examples.

abiding by the letter of the law rather than the spirit. This atmosphere creates a need for clear, consistent, rational, and enforceable rules. Penalties for violations should be tied to the cost consequences for those who suffer damage or loss as a result of the violations. The atmosphere also calls for an open rule-making process that it technically competent and especially sensitive to the desires of the groups that bear the economic and physical costs of reliability rules. Past standards were not accompanied by technical or economic analysis and justification. In the future it will be necessary to make public the analyses that justify standards on both engineering and economic grounds. Data must also be publicly available so that reliability performance can be judged and all parties can determine whether their needs are being met effectively. The rule-making process should include participation, at the board and the technical committee levels, by system operators, loads (customers), generation and transmission owners and operators, and the public.

The public interest will differ from the load's/customer's interest at times. There may be a distinct public interest in maintaining civil order, which would make avoiding geographically large outages especially important. Avoiding such outages would favor the practice of sacrificing individual loads in order to maintain overall system reliability. Promoting economic growth is another public interest that has reliability implications. Industrial loads that require reliable power also create jobs. Concerns over endangered salmon in the Pacific Northwest provide a different example where power reliability concerns conflict with a public interests concerning endangered fish. The Bonneville Power Administration declared a power emergency and reduced fish spills (water releases through dams to help salmon fingerlings swim safely to the ocean) in the spring of 2001 based on a forecast of power deficits for the following winter. The deficit forecast was in turn based partially on reliability standards for loss-of-load and reserve requirements. These reliability standards are not formally or publicly developed, however. The Northwest Power Planning Council, which is charged with ensuring a reliable power supply for the region while also protecting the environment, is encouraging the region to formally agree on reliability standards to help in the public process of balancing energy and environmental needs (Fazio 2001).

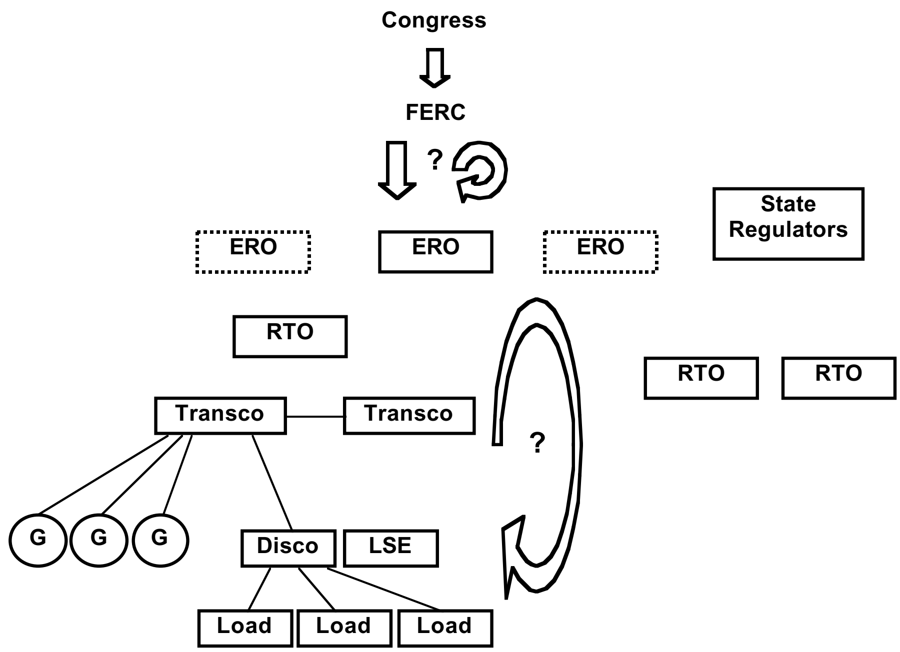
Scope and Authority

The questions of scope and authority for enforcing reliability rules are intertwined. There is general recognition that voluntary compliance with best-practice guidelines will not be sufficient in a competitive environment where individuals can profit by pushing risk onto others (Cook 2001). State regulatory authority is also not sufficient, especially because FERC has federal authority over interstate commerce. It appears that FERC does not now have sufficient authority to establish and enforce national reliability rules, and it certainly does not have authority to establish and enforce international rules that would cover entire interconnections.

Several bills have been introduced into the U.S. Congress to address reliability rules. The details of each proposal are in flux, so we do not debate them individually here. However, we discuss some of the basic concepts contained in the bills. As illustrated in Figure 5, all of the bills ask the U.S. Congress to give FERC authority to comprehensively address reliability issues for all electric utilities in the U.S. The bills differ in how they envision FERC exercising that authority, however, and to what extent reliability rules should allow regional diversity.

There are innumerable important details (and getting the details right is critical for any governance plan to

Figure 5: Who should participate in establishing reliability rules? Should FERC actively participate in the dialogue? What should the geographic scope be? Many questions remain concerning how reliability rules should be established and enforced.



succeed), but there seem to be four major questions associated with designing a reliability structure.

The first question concerns FERC's ongoing involvement in defining reliability rules. All of the proposals pass congressional authority through FERC. Some have FERC approve the initial formation of one or more Electric Reliability Organizations (EROs) and then allow FERC to delegate its authority to those organizations. Other proposals require the EROs to return to FERC for approval of each change in reliability rules. Some proposals allow FERC to take direct action to establish or modify reliability rules. Others require FERC to wait until a participant brings an appeal. The core issue is how involved FERC should be in ongoing details of reliability management. There is concern that FERC does not currently have sufficient staff with appropriate technical qualifications to take on the added function of participating actively in developing and enforcing reliability rules; it seems likely that FERC and/or EROs will require more technical staff in the future to deal with analyzing, establishing, defending, and enforcing reliability rules. Voluntary organizations may not be able to cope with the increased controversy that will likely surround reliability rules. Self-regulation works in other industries (the securities industry utilizes the National Association of Securities Dealers, for example) where federal agencies delegate power to nongovernmental entities (Michael 1993). Insuring that the EROs do not tailor rules and enforcement to serve the interests of favored parties rather than the interests of the public is critical for self-regulation to work.

The second question concerns national versus regional standards. All of the proposals before Congress recognize the physical differences among power systems in different regions of the country, and all allow differences in reliability rules when necessary. Striving for national (continental) rules and granting exceptions when necessary would reduce "seams" (differences among regions) issues and make it easier for market participants to operate in multiple regions. Allowing regions to develop reliability rules independently would mean that the rules could be tailored to the requirements and preferences of each region. The question may come down to who bears the burden of proof: Does a region have to prove that it needs a waiver from national rules? Or are all regional rules accepted as long as they are effective for insuring reliability? One possibility is to have national rules and definitions but allow region-specific determination of required reserve quantities and reliability goals/objectives.

A third question concerns how many EROs are established. If FERC establishes several EROs, then it is likely that there will be significant regional diversity. If FERC establishes a single strong ERO or if FERC itself becomes heavily involved in detailed rule development, then it is more likely that national or continental standards will emerge. Some have suggested that the regional transmission organizations (RTOs), strongly encouraged in FERC Order 2000, should also be the EROs. This is possible but raises concerns about a single organization establishing rules, performing some reliability functions, purchasing other reliability functions, judging performance, and imposing sanctions. It would be difficult for such an organization to appear to be impartial.

A fourth question concerns the involvement of the true "customers" of reliability decisions: loads (customers) and the public. These constituencies do not typically have strong technical backgrounds in transmission reliability and are also typically too small individually to participate in reliability decisions. In addition, electricity system reliability is not their primary focus. Yet they are the only reasons that the system exists, and they are the ones who pay the lion's share of the costs for both reliability and unreliability. Increasing the involvement of state regulators (as customer representatives) in developing reliability rules may be one way to address the issue.

Recommendations

DOE, FERC, and others can take a number of actions to improve the reliability of the bulk-power system:

- Promote passage of federal legislation that grants FERC authority over bulk-power reliability in the United States. FERC authority would cover all bulk-power participants, including all transmission owners (the municipal, rural cooperative, state, and federal utilities not now subject to FERC oversight as well as investor-owned utilities). In addition, FERC authority would cover all generators connected to the grid, power marketers and brokers, distribution utilities, and load-serving entities. Contracts between the system operators and each market participant should be considered substitutes for federal legislation only if federal legislation proves impossible to enact.
- Upon passage of the legislation suggested above, FERC would develop (or cause to be developed) and approve mandatory reliability standards and the associated compliance and penalty provisions required to implement such standards. That is, today's system of voluntary compliance with standards developed by a small group of industry insiders would be replaced by mandatory compliance with standards developed in an open and inclusive process.

- FERC should develop market-based penalties for failure to comply with reliability standards. That is, the penalties should be a function of the costs to the bulk-power system and to retail customers of the failure to comply. In addition, the penalties should recognize whether the failure to comply was intentional (e.g., the owner of the generator decided to sell capacity committed to contingency reserves as energy in another system) or was inadvertent (e.g., a generator suddenly tripped off line).

- To support compliance with mandatory reliability standards, FERC should develop metrics for reliability services. It is not possible to buy or sell what we cannot measure, nor is it possible to impose penalties on nonperformance with regard to something that we cannot measure. Metrics should be developed in an open, public forum and should be consistent throughout the country. There may be regional differences concerning how much of a particular reliability service is required, but the metric for assessing the quality of the service should be consistent. (The speed limit, for example, varies from road to road depending on local conditions, but the metric is consistently miles per hour.)

- FERC should conduct and publish an analysis of the benefits and costs of each reliability standard. This analysis, using historical data and simulation models, should show the pros and cons of different kinds of standards and of weaker and stricter levels for the particular standard chosen. For example, the current DCS requires that control areas recover from all disturbances within 15 minutes. Analysis could show the benefits and costs of changing the standard to 10 minutes (increased reliability, higher costs) or 20 minutes (decreased reliability, lower costs).

- FERC, DOE, and the National Association of Regulatory Utility Commissioners (NARUC) should develop and implement reporting requirements for reliability events. These requirements would provide for the collection of data that are now lacking on the number, extent, and effects of outages that interrupt service to retail customers. Separate requirements might be developed for distribution utilities and RTOs to reflect differences in distribution and bulk-power outages. These data should be made public to facilitate public choices about reliability needs and preferences.

- FERC should analyze differences among regions with respect to transmission topology, types and number of generating stations, types and magnitudes of retail load, and other factors to determine whether regional reliability standards are appropriate. This analysis should help FERC decide whether national (actually, North American) standards should be the default; if national standards prevailed, regional variances would be approved only with a clear demonstration of their value or of the need for them. If the study led FERC to decide that regional standards are preferable, national standards should be used only where regional differences are minor.

- FERC should establish compensation requirements for loads that are involuntarily curtailed. Required compensation would eliminate any incentive to use involuntary load curtailment as a resource simply because it would be cheaper than procuring adequate reserves.

- Reliability services established by FERC should be technology neutral. They should focus on the required function, not on the technology used to deliver the service. Demand-side solutions should be encouraged to complement historic generation-side solutions. Services like spinning reserve, for example, functionally involve real-power response to rebalance generation and load. The service should be defined based upon the function (real-power response within a defined time frame) not the technology (generation connected to the system).

- Control area size should be based on rational criteria. DOE should commission a study to determine why so many small control areas continue to exist and whether their numbers adversely impact system reliability by making coordination difficult or impeding commerce by increasing transaction costs. FERC should act to eliminate the incentives to operate small control areas if the study shows that they adversely affect the system.

Summary

Restructuring the electric power industry in the U.S. is dramatically affecting reliability management and oversight. The physics of the power system are not changing, but the commercial relationships and economic interests of various parties are. In a simplified power system, reliability requirements are straightforward, and reserve requirements are deterministic; however, reliability rules need to reflect the complexity of the real world, with which system operators have always had to grapple. In the face of strong, competing economic interests, reliability requirements must be defined clearly for each party. Metrics, pay for performance, and/or enforcement are required when competitive interests differ from communal ones. Metrics are developing slowly, and enforcement awaits resolution of governance issues and the establishment of a chain of authority for reliability rules. The historic voluntary structure that worked well for the vertically integrated utilities of the past is not adequate today.

Load has always been used as a reliability resource, at least in the last extreme and generally without consent. A market structure should be fostered in which loads (customers) could voluntarily respond to reliability needs and be compensated for their contributions. A competitive market could set the value of the contributions.

Managing reliability is managing risk. The unique features of AC electric power (the passive nature of the transmission system coupled with the need to continuously balance load with generation) result in a communal power system that exposes all users to the shared risk of system collapse. It is not practical to build a power system that is 100 percent reliable. Reliability rules establish how much risk the system will assume. Deciding how much risk to take and selecting reliability rules should be communal decisions.

Two major questions associated with reliability management are who decides on the acceptable level of risk (and the costs to maintain that level of reliability) and who takes the risk (and incurs the cost). Societal as well as individual interests must be considered.

Acknowledgments

We thank Terry Boston, David Meyer, Dave Nevius, Richard Sedano, and Alison Silverstein for their very helpful comments on a draft of this paper.

References

California ISO. 2001. California ISO System Conditions Notification Process. Folsom CA, June 15.

Cook, D. 2001. Hearing Before the United States House of Representatives Subcommittee on Energy and Air Quality, Prepared Testimony. Princeton, N.J.: North American Electric Reliability Council, October 10.

Davis, T. 2001. "Davis, Cal-ISO Struggle to Make Power Purchase Strategy Work." *The Energy Daily,* vol. 29, no. 204, October 23.

Dintelman, R. 2001. WSCC Reliability Management System (RMS) Update. EPITF Meeting. San Francisco CA, June 26.

Fazio, J. 2001. *Reliability Standards for the Northwest Power Planning Council.* Portland OR: Northwest Power Planning Council.

Florida Reliability Coordinating Council (FRCC). 1999. *FRCC Standards Handbook.* Tampa FL, October.

Hirst, E. and B. Kirby. 2001a. "Retail-Load Participation" in *Competitive Wholesale Electricity Markets.* Washington, D.C.: Edison Electric Institute, and Alexandria, VA: Project for Sustainable FERC Energy Policy, January.

Hirst, E. and B. Kirby. 2001b. "Measuring Generator Performance" in *Providing Regulation and Load-Following Ancillary Services.* ORNL/TM-2000-383, Oak Ridge TN: Oak Ridge National Laboratory, January.

ISO New England, Inc. 2001. *System Operations Procedures.* Holyoke MA, July 18.

Michael, D. 1993. Federal Agency Use of Audited Self-Regulation as a Regulatory Technique. Washington D.C.: American University Washington College of Law, *Administrative Law Review,* vol. 47, page 171, November.

NERC (North American Electric Reliability Council). 2001a. *NERC Operating Manual.* Princeton NJ, 12 July.

NERC. 2001b. *Roster.* Princeton NJ, October.

Smith, J. 2001. Arizona Corporation Commission staff. Statement at the U.S. DOE Transmission Grid Study Workshop, Phoenix Arizona, September 28.

Western Systems Coordinating Council (WSCC) . 2001. Reliability Criteria Agreement, November. download from www.wscc.com